SUPPLEMENTARY MATERIAL

A structural enriched functional network: an application to predict brain cognitive performance

ManSu Kim, Jingxuan Bao, Kefei Liu, Bo-yong Park, Hyunjin Park, Jae Young Baik, Li Shen

1. Optimization algorithm for Eq. (8)

The objective function in Eq. (3) is a convex optimization problem with differentiable objective and constraint functions and is strictly feasible (Slater's condition holds), the Karush–Kuhn–Tucker conditions provide necessary and sufficient conditions for optimality (Boyd et al., 2004). We can write the Lagrangian function of the Eq. (3) as:

$$\frac{1}{2}\left\|\boldsymbol{\beta}_i - \boldsymbol{\alpha}_i^{(k-1)}\right\|_2^2 - \gamma\left(\boldsymbol{\beta}_i^{\mathrm{T}}\mathbf{1} - 1\right) - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\beta}_i, \tag{8}$$

where γ and $\boldsymbol{\lambda}$ are a Lagrangian multiplier and Lagrangian multiplier vector, respectively and both of them are to be determined. Suppose that the optimal solution to the proximal problem (3) is $\boldsymbol{\beta}^*$, the associated Lagrangian multipliers are γ^* and $\boldsymbol{\lambda}^*$.

$$\forall j,\ \boldsymbol{\beta}_{i\,j}^* - \boldsymbol{\alpha}_i^{(k-1)}{}_j - \gamma^* - \boldsymbol{\lambda}_j^* = 0, \tag{9}$$

$$\forall j, \boldsymbol{\beta}_{i\,j}^* \geq 0, \tag{10}$$

$$\forall j, \boldsymbol{\lambda}_j^* \geq 0, \tag{11}$$

$$\forall j, \boldsymbol{\beta}_{i\,j}^* \boldsymbol{\lambda}_j^* = 0, \tag{12}$$

where, $\boldsymbol{\beta}_{i\,j}^*$ denoted the j-th element of vector $\boldsymbol{\beta}_i^*$. We can rewrite Eq. (9) as $\boldsymbol{\beta}_{i\,j}^* - \boldsymbol{\alpha}_i^{(k-1)}{}_j - \gamma^* 1 - \boldsymbol{\lambda}_j^* = 0$. We have $\gamma^* = \frac{1 - \mathbf{1}^T\boldsymbol{\alpha}_i^{(k-1)} - \mathbf{1}^T\boldsymbol{\lambda}^*}{n}$ using the constraint $\boldsymbol{\beta}_i^T\mathbf{1} = 1$ and derive $\boldsymbol{\beta}^* = \left(\boldsymbol{v} - \frac{1}{n}\mathbf{1}^T\boldsymbol{\alpha}_i^{(k-1)} + \frac{1}{n}\mathbf{1} - \frac{\mathbf{1}^T\boldsymbol{\lambda}^*}{n}\mathbf{1}\right) + \boldsymbol{\lambda}^*$. We rewrite it as $\boldsymbol{\beta}^* = \boldsymbol{u} + \boldsymbol{\lambda}^* - \bar{\lambda}^*\mathbf{1}$, where $\bar{\lambda}^* = \frac{\mathbf{1}^T\boldsymbol{\lambda}^*}{n}$ and $\boldsymbol{u} = \boldsymbol{\alpha}_i^{(k-1)} - \frac{1}{n}\mathbf{1}^T\boldsymbol{\alpha}_i^{(k-1)} + \frac{1}{n}\mathbf{1}$. Thus, $\forall j$ we have

$$\boldsymbol{\beta}_{i\,j}^* = \boldsymbol{u}_j + \boldsymbol{\lambda}_j^* - \bar{\lambda}^*. \tag{13}$$

According to Eqs. (10)-(13), we have $\boldsymbol{u}_j + \boldsymbol{\lambda}_j^* - \bar{\lambda}^* = \left(\boldsymbol{u}_j - \bar{\lambda}^*\right)_+$, where $x_+ = max(x, 0)$. We then have $\boldsymbol{\beta}_{i\,j}^* = \left(\boldsymbol{u}_j - \bar{\lambda}^*\right)_+$. Therefore, given we know $\bar{\boldsymbol{\lambda}}^*$, we can compute the optimal solution $\boldsymbol{\beta}^*$.

To obtain $\bar{\boldsymbol{\lambda}}^*$, we rewrite Eq. (13) as $\boldsymbol{\lambda}_j^* = \bar{\lambda}^* + \boldsymbol{\beta}_{i\,j}^* - \boldsymbol{u}_j$. According to Eqs. (10)-(13), we have $\boldsymbol{\lambda}_j^* = \left(\bar{\lambda}^* - \boldsymbol{u}_j\right)_+$. Since $\boldsymbol{v}$ is a $n-1$ dimensional vector, we have $\bar{\lambda}^* = \frac{1}{n-1}\sum_{j=1}^{n-1}\left(\bar{\lambda}^* - \boldsymbol{u}_j\right)_+$. Therefore, we define a function as follows:

$$f\left(\bar{\lambda}^*\right) = \frac{1}{n-1}\sum_{j=1}^{n-1}\left(\bar{\lambda}^* - \boldsymbol{u}_j\right)_+ - \bar{\lambda}, \tag{14}$$

and we obtain $\bar{\lambda}^*$ by solving Eq. (14) to be zero. Since $\lambda^* \geq 0, f'\left(\bar{\lambda}^*\right) \leq 0,$ and $f'\left(\bar{\lambda}^*\right)$ is a piecewise linear and convex function, we can compute the root of $f'\left(\bar{\lambda}^*\right) = 0$ using the Newton method efficiently.

2. Comparison of predictive performance with four different regional measures in the proposed model

We performed additional analysis to predict outcomes with four other regional measures, including betweenness centrality, eigenvector centrality, clustering coefficients, and local efficiency, computed from the proposed model. We found that clustering coefficients and local efficiency showed similar performances compared with degree centrality (DC) in terms of the correlation coefficient, as shown in Table S2. Recent studies reported that DC has been widely used to examine nodal characteristics of brain network (Buckner et al., 2009; Bullmore & Sporns, 2009; Cole et al., 2010; He et al., 2009). A study reported that eigenvector centrality and page-rank centrality showed higher centrality compared with DC for subcortical regions due to their own self connection (Zuo et al., 2012). Thus, we focused on the study of DC as the centrality measure to predict the clinical outcomes in the main text.

3. Supplementary Figures

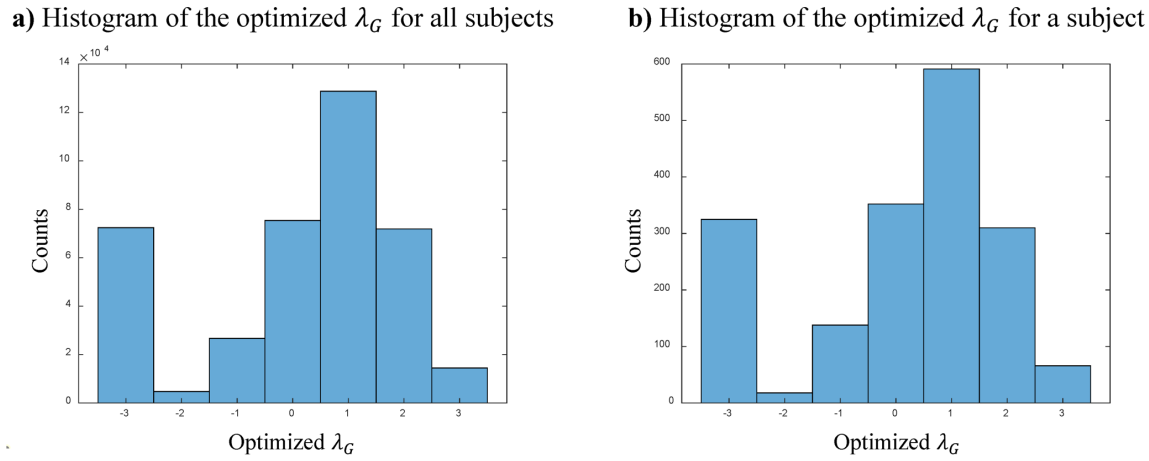

Figure S1. The histogram of the optimized λ_G during the 5-fold cross-validation tunning process. The sub-figures (a) and (b) showed the distributions of the optimized λ_G for all subjects and for one specific subject, respectively. For histogram plots, we used a base 10 logarithmic scale for the histogram bins.

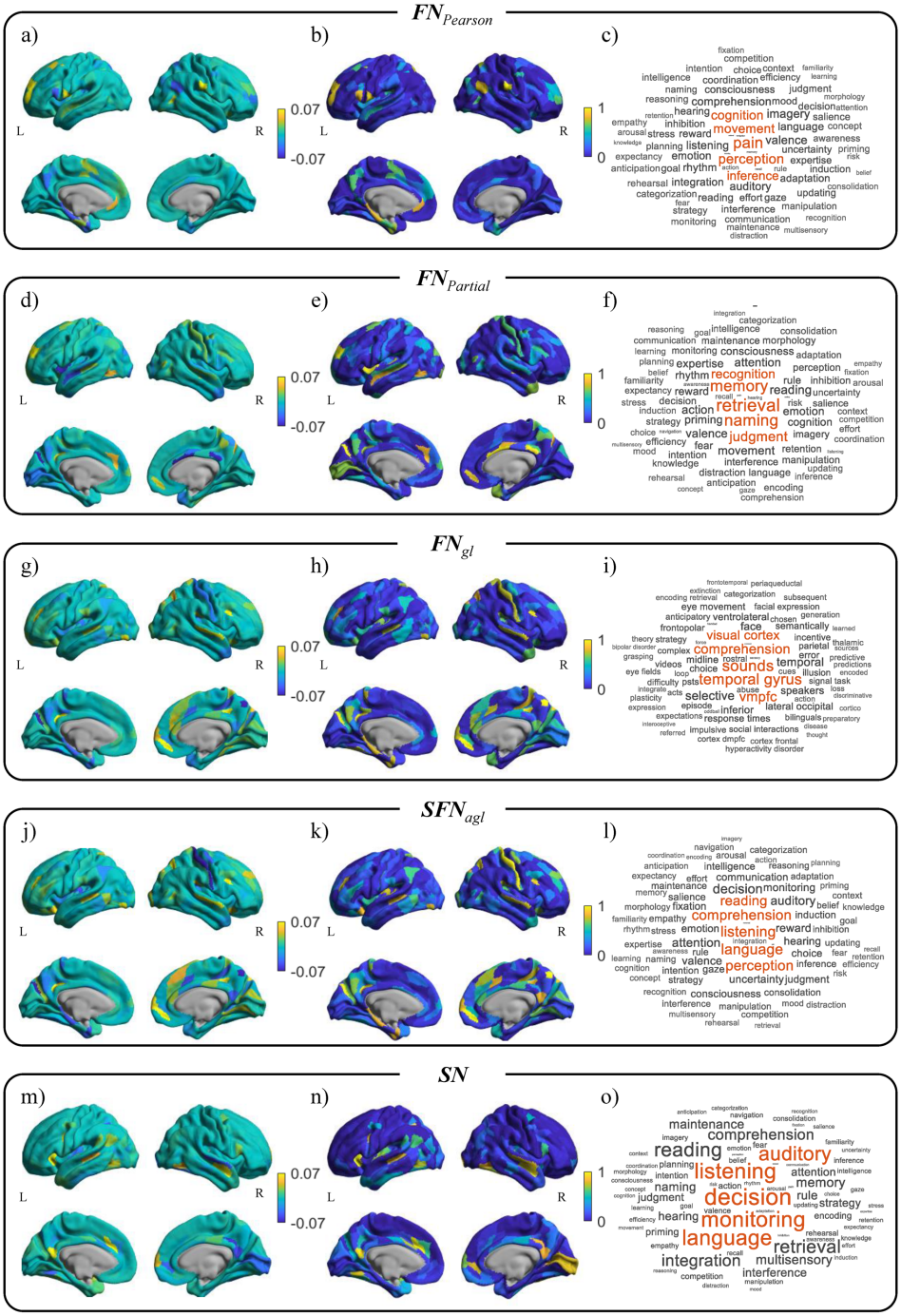

Figure S2. The activation pattern maps and the related cognitive topics for predicting working memory 2-back accuracy. The mean standardized regression coefficients and the decoding results for six different network models were plotted in each row. For each row, left column: mean standardized regression coefficients map. Center column: selection probability map. Right column: word clouds plot related to cognitive function in the Neurosynth database.

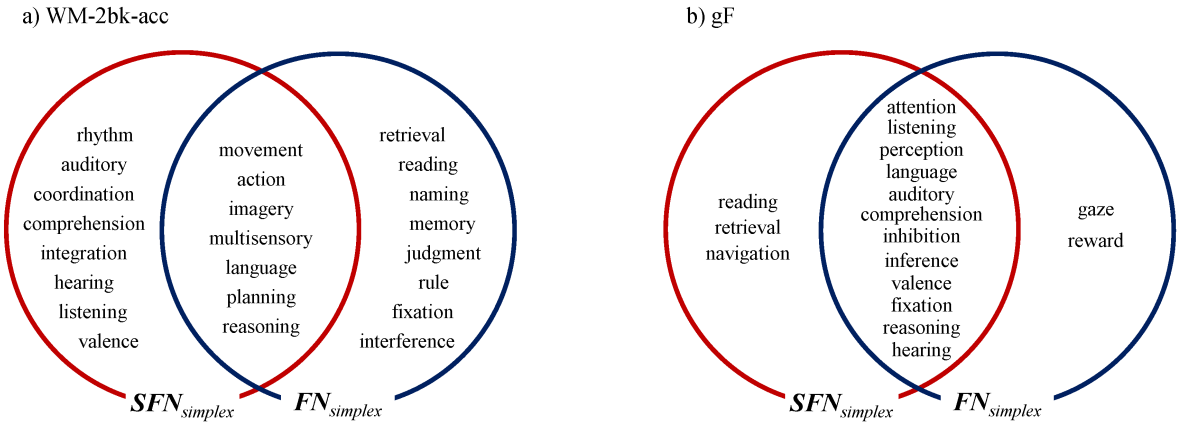

Figure S3. Venn diagrams for Neurosynth topics contributing to the prediction model based on *SFN*$_{simplex}$ and *FN*$_{simplex}$.

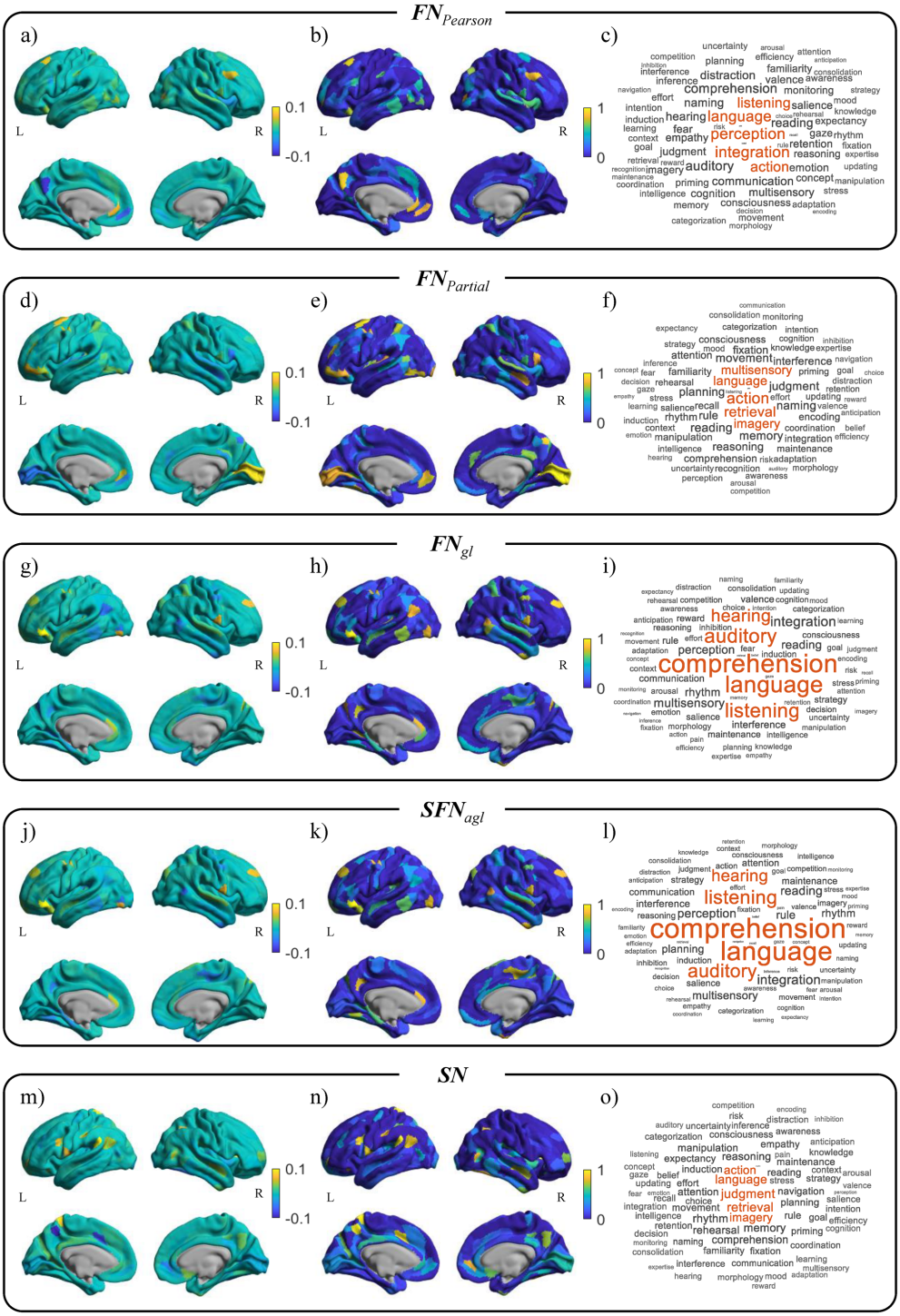

Figure S4. The activation pattern map and the related cognitive topics for predicting fluid intelligence. The mean standardized regression coefficients and the decoding results for six different network models were plotted in each row. For each row, left column: mean standardized regression coefficients map. Center column: selection probability map. Right column: word clouds plot related to cognitive function in the Neurosynth database.

4. Supplementary Tables

Table S1. Comparison of Neurosynth topics contributing to the prediction model based on $SFN_{simplex}$ and $FN_{simplex}$. The common topics between $SFN_{simplex}$ and $FN_{simplex}$ were marked with an asterisk.

Rank	*WM-2bk-acc*		*gF*	
	$SFN_{simplex}$	$FN_{simplex}$	$SFN_{simplex}$	$FN_{simplex}$
1	comprehension*	listening*	movement*	action*
2	listening*	comprehension*	action*	retrieval
3	perception*	attention*	imagery*	language*
4	language*	auditory*	rhythm	imagery*
5	auditory*	perception*	auditory	multisensory*
6	attention*	hearing*	multisensory*	reading
7	hearing*	language*	coordination	planning*
8	reading	fixation*	comprehension	naming
9	retrieval	inference*	language*	memory
10	inference*	reasoning*	integration	movement*
11	valence*	decision	planning*	judgment
12	fixation*	inhibition*	hearing	rule
13	reasoning*	gaze	reasoning*	reasoning*
14	navigation	reward	listening	fixation
15	inhibition*	valence*	valence	interference

Table S2. Comparison of the prediction models based on $SFN_{simplex}$ to predict WM-2bk-acc and gF. The prediction performance was reported in terms of RMSE and correlation coefficient (*r*) between actual and predicted scores. The graph metric acronyms were reported as follows: DC=degree centrality, BC=Betweenness centrality, EVC=eigenvector centrality, CC=Clustering coefficient, and Eloc=local efficiency.

		DC	**BC**	**EVC**	**CC**	**Eloc**
WM-2bk-acc	*r*	0.304	0.107	0.181	0.291	0.221
	RMSE	10.47	10.52	10.51	10.30	10.33
gF	*r*	0.287	0.183	-0.026	0.219	0.267
	RMSE	4.54	4.20	4.27	4.16	4.11